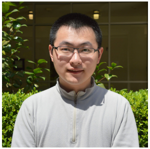
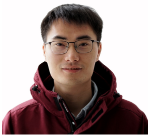
No-Regret Learning in Time-Varying Zero-Sum Games

Haipeng Luo[1]

joint work w. Mengxiao Zhang[1], Peng Zhao[2], and Zhi-Hua Zhou[2]


1. University of Southern California
2. Nanjing University

Introduction

Uncoupled learning dynamics for a fixed game is well studied.

Introduction

Uncoupled learning dynamics for a fixed game is well studied.

What if the game is changing?

Introduction

Uncoupled learning dynamics for a fixed game is well studied.

What if the game is changing?

- in some cases, changes depend on players' actions (Markov games)

Introduction

Uncoupled learning dynamics for a fixed game is well studied.

What if the game is changing?

- in some cases, changes depend on players' actions (Markov games)

An example from Éva Tardos' talk:

Second-by-second packet traffic

(changes come from unserved packets, determined by players)

Introduction

Uncoupled learning dynamics for a fixed game is well studied.

What if the game is changing?

- in some cases, changes depend on players' actions (Markov games)
- in other cases, changes are only caused by exogenous factors

An example from Éva Tardos' talk:

Second-by-second packet traffic

(changes come from unserved packets, determined by players)

Introduction

Uncoupled learning dynamics for a fixed game is well studied.

What if the game is changing?

- in some cases, changes depend on players' actions (Markov games)
- in other cases, changes are only caused by exogenous factors

An example from Éva Tardos' talk:

Second-by-second packet traffic

Morning rush-hour traffic

(changes come from unserved packets, determined by players)

(changes come from road constructions or accidents, not determined by players)

This Talk

Focus: uncoupled learning over a sequence of time-varying zero-sum normal-form games *decided exogenously by the environment*.

This Talk

Focus: uncoupled learning over a sequence of time-varying zero-sum normal-form games *decided exogenously by the environment*.

First part: how to measure performance?

This Talk

Focus: uncoupled learning over a sequence of time-varying zero-sum normal-form games *decided exogenously by the environment*.

First part: how to measure performance?

- review an existing measure (and argue why it is problematic)

This Talk

Focus: uncoupled learning over a sequence of time-varying zero-sum normal-form games *decided exogenously by the environment*.

First part: how to measure performance?

- review an existing measure (and argue why it is problematic)
- consider/propose three natural measures

This Talk

Focus: uncoupled learning over a sequence of time-varying zero-sum normal-form games *decided exogenously by the environment*.

First part: how to measure performance?

- review an existing measure (and argue why it is problematic)
- consider/propose three natural measures

Second part: propose one single algorithm that

This Talk

Focus: uncoupled learning over a sequence of time-varying zero-sum normal-form games *decided exogenously by the environment*.

First part: how to measure performance?

- review an existing measure (and argue why it is problematic)
- consider/propose three natural measures

Second part: propose one single algorithm that

- is parameter-free

This Talk

Focus: uncoupled learning over a sequence of time-varying zero-sum normal-form games *decided exogenously by the environment*.

First part: how to measure performance?

- review an existing measure (and argue why it is problematic)
- consider/propose three natural measures

Second part: propose one single algorithm that

- is parameter-free
- achieves strong guarantees under all three measures

This Talk

Focus: uncoupled learning over a sequence of time-varying zero-sum normal-form games *decided exogenously by the environment*.

First part: how to measure performance?

- review an existing measure (and argue why it is problematic)
- consider/propose three natural measures

Second part: propose one single algorithm that

- is parameter-free
- achieves strong guarantees under all three measures
- recovers best known results when the game is fixed

Setup

Time-Varying Zero-Sum Games

For each round $t = 1, \ldots, T$:

Time-Varying Zero-Sum Games

For each round $t = 1, \ldots, T$:

- environment decides a payoff matrix $A_t \in [-1, 1]^{m \times n}$;

Time-Varying Zero-Sum Games

For each round $t = 1, \ldots, T$:

- environment decides a payoff matrix $A_t \in [-1, 1]^{m \times n}$;
- without knowing A_t, x-player decides a mixed strategy $x_t \in \Delta_m$ and y-player decides a mixed strategy $y_t \in \Delta_n$;

Time-Varying Zero-Sum Games

For each round $t = 1, \ldots, T$:

- environment decides a payoff matrix $A_t \in [-1, 1]^{m \times n}$;
- without knowing A_t, x-player decides a mixed strategy $x_t \in \Delta_m$ and y-player decides a mixed strategy $y_t \in \Delta_n$;
- x-player suffers loss $x_t^\top A_t y_t$ and observes $A_t y_t$, while y-player receives reward $x_t^\top A_t y_t$ and observes $x_t^\top A_t$ (mixture feedback).

Time-Varying Zero-Sum Games

For each round $t = 1, \ldots, T$:

- environment decides a payoff matrix $A_t \in [-1, 1]^{m \times n}$;
- without knowing A_t, x-player decides a mixed strategy $x_t \in \Delta_m$ and y-player decides a mixed strategy $y_t \in \Delta_n$;
- x-player suffers loss $x_t^\top A_t y_t$ and observes $A_t y_t$, while y-player receives reward $x_t^\top A_t y_t$ and observes $x_t^\top A_t$ (mixture feedback).

More applications:

- online linear programming (Agrawal-Wang-Ye'14)
- adversarial bandits w. knapsacks (Immorlica-Sankararaman-Schapire-Slivkins'18)

Related Work

Cardoso-Abernethy-Wang-Xu'19 (most related):

- their feedback is either the entire matrix A_t, or one entry sampled from (x_t, y_t)

Related Work

Cardoso-Abernethy-Wang-Xu'19 (most related):

- their feedback is either the entire matrix A_t, or one entry sampled from (x_t, y_t)

Roy-Chen-Balasubramanian-Mohapatra'19:

- time-varying convex-concave games

Related Work

Cardoso-Abernethy-Wang-Xu'19 (most related):

- their feedback is either the entire matrix A_t, or one entry sampled from (x_t, y_t)

Roy-Chen-Balasubramanian-Mohapatra'19:

- time-varying convex-concave games

Duvocelle-Mertikopoulos-Staudigl-Vermeulen'21:

- time-varying strongly monotone games

Related Work

Cardoso-Abernethy-Wang-Xu'19 (most related):

- their feedback is either the entire matrix A_t, or one entry sampled from (x_t, y_t)

Roy-Chen-Balasubramanian-Mohapatra'19:

- time-varying convex-concave games

Duvocelle-Mertikopoulos-Staudigl-Vermeulen'21:

- time-varying strongly monotone games

Fiez-Sim-Skoulakis-Piliouras-Ratliff'21:

- periodic zero-sum games

First Part: How to Measure Performance?

Classical Individual Regret

Time-varying A_t or not, it always makes sense to selfishly minimize regret:

$$\mathrm{Reg}_T^x = \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \sum_{t=1}^{T} x^\top A_t y_t$$

Classical Individual Regret

Time-varying A_t or not, it always makes sense to selfishly minimize regret:

$$\mathrm{Reg}_T^x = \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \sum_{t=1}^{T} x^\top A_t y_t$$

$$\mathrm{Reg}_T^y = \max_{y \in \Delta_n} \sum_{t=1}^{T} x_t^\top A_t y - \sum_{t=1}^{T} x_t^\top A_t y_t$$

Beyond Individual Regret

Any game-theoretic guarantees?

Beyond Individual Regret

Any game-theoretic guarantees?

- when $A_t = A$ is fixed, can argue closeness of $\left(\frac{1}{T}\sum_{t=1}^{T} x_t, \frac{1}{T}\sum_{t=1}^{T} y_t\right)$ or even (x_T, y_T) to the Nash Equilibria of A.

Beyond Individual Regret

Any game-theoretic guarantees?

- when $A_t = A$ is fixed, can argue closeness of $\left(\frac{1}{T}\sum_{t=1}^T x_t, \frac{1}{T}\sum_{t=1}^T y_t\right)$ or even (x_T, y_T) to the Nash Equilibria of A.
- what can we say when A_t is changing over time?

One Proposal: NE-Regret

Cardoso-Abernethy-Wang-Xu'19 proposes Nash Equilibrium regret:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right|$$

One Proposal: NE-Regret

Cardoso-Abernethy-Wang-Xu'19 proposes Nash Equilibrium regret:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right|$$

Issues:

- incompatible with classical regret!

One Proposal: NE-Regret

Cardoso-Abernethy-Wang-Xu'19 proposes Nash Equilibrium regret:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right|$$

Issues:

- incompatible with classical regret! No algorithm can achieve $o(T)$ bounds for Reg_T^x, Reg_T^y, and NE-Reg_T simultaneously [CAWX'19].

One Proposal: NE-Regret

Cardoso-Abernethy-Wang-Xu'19 proposes Nash Equilibrium regret:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right|$$

Issues:

- incompatible with classical regret! No algorithm can achieve $o(T)$ bounds for Reg_T^x, Reg_T^y, and NE-Reg_T simultaneously [CAWX'19].
- we show that it can be $\Omega(T)$ even for "perfect" players!

One Proposal: NE-Regret

Cardoso-Abernethy-Wang-Xu'19 proposes Nash Equilibrium regret:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right|$$

Issues:

- incompatible with classical regret! No algorithm can achieve $o(T)$ bounds for Reg_T^x, Reg_T^y, and NE-Reg_T simultaneously [CAWX'19].
- we show that it can be $\Omega(T)$ even for "perfect" players!
  Consider: $A_t = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ if $t \leq T/2$ or $\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$ otherwise;

One Proposal: NE-Regret

Cardoso-Abernethy-Wang-Xu'19 proposes Nash Equilibrium regret:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right|$$

Issues:

- incompatible with classical regret! No algorithm can achieve $o(T)$ bounds for Reg_T^x, Reg_T^y, and NE-Reg_T simultaneously [CAWX'19].
- we show that it can be $\Omega(T)$ even for "perfect" players!
  Consider: $A_t = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ if $t \leq T/2$ or $\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$ otherwise; even if both players alway play the Nash, we have

$$\text{NE-Reg}_T = \left| \frac{T}{2} - \min_{x \in \Delta_m} \max_{y \in \Delta_n} x^\top \begin{pmatrix} T & -T \\ 0 & 0 \end{pmatrix} y \right| = \left| \frac{T}{2} - 0 \right| = \frac{T}{2}$$

Our Proposal: Dynamic NE-Regret

We propose to move the "minmax" inside the summation:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right| \qquad \text{(CAWX'19)}$$

$$\text{DynNE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \sum_{t=1}^{T} \min_{x \in \Delta_m} \max_{y \in \Delta_n} x^\top A_t y \right| \qquad \text{(Ours)}$$

Our Proposal: Dynamic NE-Regret

We propose to move the "minmax" inside the summation:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right| \qquad \text{(CAWX'19)}$$

$$\text{DynNE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \sum_{t=1}^{T} \min_{x \in \Delta_m} \max_{y \in \Delta_n} x^\top A_t y \right| \qquad \text{(Ours)}$$

- the connection is (on the surface) analogous to the classical regret and its dynamic version

$$\text{Reg}_T^x = \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \sum_{t=1}^{T} x^\top A_t y_t$$

$$\text{DynReg}_T^x = \sum_{t=1}^{T} x_t^\top A_t y_t - \sum_{t=1}^{T} \min_{x \in \Delta_m} x^\top A_t y_t \qquad \text{(Zinkevich'03)}$$

Our Proposal: Dynamic NE-Regret

We propose to move the "minmax" inside the summation:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right| \qquad \text{(CAWX'19)}$$

$$\text{DynNE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \sum_{t=1}^{T} \min_{x \in \Delta_m} \max_{y \in \Delta_n} x^\top A_t y \right| \qquad \text{(Ours)}$$

- the connection is (on the surface) analogous to the classical regret and its dynamic version
- but while $\text{Reg}_T^x \leq \text{DynReg}_T^x$, DynNE-Reg_T could be smaller than NE-Reg_T

Our Proposal: Dynamic NE-Regret

We propose to move the "minmax" inside the summation:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x\in\Delta_m} \max_{y\in\Delta_n} \sum_{t=1}^{T} x^\top A_t y \right| \qquad \text{(CAWX'19)}$$

$$\text{DynNE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \sum_{t=1}^{T} \min_{x\in\Delta_m} \max_{y\in\Delta_n} x^\top A_t y \right| \qquad \text{(Ours)}$$

- the connection is (on the surface) analogous to the classical regret and its dynamic version
- but while $\text{Reg}_T^x \le \text{DynReg}_T^x$, DynNE-Reg_T could be smaller than NE-Reg_T (e.g. $\text{DynNE-Reg}_T = 0$ if both players alway play Nash)

Our Proposal: Dynamic NE-Regret

We propose to move the "minmax" inside the summation:

$$\text{NE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \min_{x \in \Delta_m} \max_{y \in \Delta_n} \sum_{t=1}^{T} x^\top A_t y \right| \qquad \text{(CAWX'19)}$$

$$\text{DynNE-Reg}_T = \left| \sum_{t=1}^{T} x_t^\top A_t y_t - \sum_{t=1}^{T} \min_{x \in \Delta_m} \max_{y \in \Delta_n} x^\top A_t y \right| \qquad \text{(Ours)}$$

- the connection is (on the surface) analogous to the classical regret and its dynamic version
- but while $\text{Reg}_T^x \le \text{DynReg}_T^x$, DynNE-Reg_T could be smaller than NE-Reg_T (e.g. $\text{DynNE-Reg}_T = 0$ if both players alway play Nash)
- more importantly, DynNE-Reg_T is compatible with Reg_T^x (as we will see)

Duality Gap

Approaching the minimax values does not imply closeness to Nash.

Duality Gap

Approaching the minimax values does not imply closeness to Nash.

Thus, another natural measure is the (cumulative) duality gap:

$$\text{Dual-Gap}_T = \sum_{t=1}^{T} \left(\max_{y \in \Delta_n} x_t^\top A_t y - \min_{x \in \Delta_m} x^\top A_t y_t \right)$$

Duality Gap

Approaching the minimax values does not imply closeness to Nash.

Thus, another natural measure is the (cumulative) duality gap:

$$\text{Dual-Gap}_T = \sum_{t=1}^{T} \left(\max_{y \in \Delta_n} x_t^\top A_t y - \min_{x \in \Delta_m} x^\top A_t y_t \right)$$

Other possibilities: cumulative ℓ_1 distance to Nash (e.g. [RCBM'19])

Duality Gap

Approaching the minimax values does not imply closeness to Nash.

Thus, another natural measure is the (cumulative) duality gap:

$$\text{Dual-Gap}_T = \sum_{t=1}^{T} \left(\max_{y \in \Delta_n} x_t^\top A_t y - \min_{x \in \Delta_m} x^\top A_t y_t \right)$$

Other possibilities: cumulative ℓ_1 distance to Nash (e.g. [RCBM'19])

- usually depends on problem-specific constants other than n and m.

Quick Summary on Performance Measures

We consider the following three measures:

- $\mathrm{Reg}_T^x = \sum_{t=1}^T x_t^\top A_t y_t - \min_{x \in \Delta_m} \sum_{t=1}^T x^\top A_t y_t$
- $\text{DynNE-Reg}_T = \left| \sum_{t=1}^T x_t^\top A_t y_t - \sum_{t=1}^T \min_{x \in \Delta_m} \max_{y \in \Delta_n} x^\top A_t y \right|$
- $\text{Dual-Gap}_T = \sum_{t=1}^T \left(\max_{y \in \Delta_n} x_t^\top A_t y - \min_{x \in \Delta_m} x^\top A_t y_t \right)$

Quick Summary on Performance Measures

We consider the following three measures:

- $\text{Reg}_T^x = \sum_{t=1}^T x_t^\top A_t y_t - \min_{x \in \Delta_m} \sum_{t=1}^T x^\top A_t y_t$
- $\text{DynNE-Reg}_T = \left| \sum_{t=1}^T x_t^\top A_t y_t - \sum_{t=1}^T \min_{x \in \Delta_m} \max_{y \in \Delta_n} x^\top A_t y \right|$
- $\text{Dual-Gap}_T = \sum_{t=1}^T \left(\max_{y \in \Delta_n} x_t^\top A_t y - \min_{x \in \Delta_m} x^\top A_t y_t \right)$

A quick remark: one can show

$$\max \left\{ \text{Reg}_T^x, \text{Reg}_T^y, \text{DynNE-Reg}_T \right\} \le \text{Dual-Gap}_T,$$

but this upper bound can be quite *loose*.

Non-Stationarity Measures

No meaningful guarantees can be obtained (on DynNE-Reg_T and Dual-Gap_T) if the game changes arbitrarily.

Non-Stationarity Measures

No meaningful guarantees can be obtained (on DynNE-Reg_T and Dual-Gap_T) if the game changes arbitrarily.

To capture the difficulty of the problem, define non-stationarity measures:

Non-Stationarity Measures

No meaningful guarantees can be obtained (on $\mathrm{DynNE\text{-}Reg}_T$ and $\mathrm{Dual\text{-}Gap}_T$) if the game changes arbitrarily.

To capture the difficulty of the problem, define non-stationarity measures:

- variation/variance of games (with $\bar{A} = \frac{1}{T}\sum_{t=1}^{T} A_t$):

$$V_T = \sum_{t=2}^{T} \|A_t - A_{t-1}\|_\infty^2, \quad W_T = \sum_{t=1}^{T} \|A_t - \bar{A}\|_\infty,$$

Non-Stationarity Measures

No meaningful guarantees can be obtained (on DynNE-Reg_T and Dual-Gap_T) if the game changes arbitrarily.

To capture the difficulty of the problem, define non-stationarity measures:

- variation/variance of games (with $\bar{A} = \frac{1}{T}\sum_{t=1}^{T} A_t$):

$$V_T = \sum_{t=2}^{T} \|A_t - A_{t-1}\|_\infty^2, \quad W_T = \sum_{t=1}^{T} \|A_t - \bar{A}\|_\infty,$$

- variation of Nash Equilibria:

$$P_T = \min_{\forall t, (x_t^*, y_t^*) \in \text{NE of } A_t} \sum_{t=2}^{T} \left(\|x_t^* - x_{t-1}^*\|_1 + \|y_t^* - y_{t-1}^*\|_1\right),$$

Non-Stationarity Measures

No meaningful guarantees can be obtained (on $\mathrm{DynNE\text{-}Reg}_T$ and $\mathrm{Dual\text{-}Gap}_T$) if the game changes arbitrarily.

To capture the difficulty of the problem, define non-stationarity measures:

- variation/variance of games (with $\bar{A} = \frac{1}{T}\sum_{t=1}^{T} A_t$):

$$V_T = \sum_{t=2}^{T} \|A_t - A_{t-1}\|_\infty^2, \quad W_T = \sum_{t=1}^{T} \|A_t - \bar{A}\|_\infty,$$

- variation of Nash Equilibria:

$$P_T = \min_{\forall t, (x_t^*, y_t^*) \in \text{NE of } A_t} \sum_{t=2}^{T} \left(\|x_t^* - x_{t-1}^*\|_1 + \|y_t^* - y_{t-1}^*\|_1 \right),$$

- $V_t \leq 4W_T$ holds always, but they are incomparable with P_T.

Non-Stationarity Measures

No meaningful guarantees can be obtained (on $\mathrm{DynNE\text{-}Reg}_T$ and $\mathrm{Dual\text{-}Gap}_T$) if the game changes arbitrarily.

To capture the difficulty of the problem, define non-stationarity measures:

- variation/variance of games (with $\bar{A} = \frac{1}{T}\sum_{t=1}^{T} A_t$):

$$V_T = \sum_{t=2}^{T} \|A_t - A_{t-1}\|_\infty^2, \quad W_T = \sum_{t=1}^{T} \|A_t - \bar{A}\|_\infty,$$

- variation of Nash Equilibria:

$$P_T = \min_{\forall t, (x_t^*, y_t^*) \in \text{NE of } A_t} \sum_{t=2}^{T} \left(\|x_t^* - x_{t-1}^*\|_1 + \|y_t^* - y_{t-1}^*\|_1 \right),$$

- $V_t \leq 4W_T$ holds always, but they are incomparable with P_T.
- Goal: whenever (some of) V_T, W_T, P_T are $o(T)$, obtain $o(T)$ bounds for $\mathrm{DynNE\text{-}Reg}_T$ and $\mathrm{Dual\text{-}Gap}_T$

Second Part: Our Algorithm and Guarantees

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Measure	Time-Varying Game	Fixed Game
Individual Regret		
Dynamic NE-Reg		
Duality Gap		

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Measure	Time-Varying Game	Fixed Game
Individual Regret	$\widetilde{\mathcal{O}}(\sqrt{1+Q_T})$	
Dynamic NE-Reg		
Duality Gap		

- $Q_T = V_T + \min\{P_T, W_T\}$

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Measure	Time-Varying Game	Fixed Game
Individual Regret	$\widetilde{\mathcal{O}}(\sqrt{1+Q_T})$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Dynamic NE-Reg		
Duality Gap		

- $Q_T = V_T + \min\{P_T, W_T\}$

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Measure	Time-Varying Game	Fixed Game
Individual Regret	$\widetilde{\mathcal{O}}(\sqrt{1+Q_T})$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Dynamic NE-Reg	$\widetilde{\mathcal{O}}\big(\min\{\sqrt{(1+V_T)(1+P_T)}+P_T, 1+W_T\}\big)$	
Duality Gap		

- $Q_T = V_T + \min\{P_T, W_T\}$

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Measure	Time-Varying Game	Fixed Game
Individual Regret	$\widetilde{\mathcal{O}}(\sqrt{1+Q_T})$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Dynamic NE-Reg	$\widetilde{\mathcal{O}}\left(\min\{\sqrt{(1+V_T)(1+P_T)}+P_T, 1+W_T\}\right)$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Duality Gap		

- $Q_T = V_T + \min\{P_T, W_T\}$

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Measure	Time-Varying Game	Fixed Game
Individual Regret	$\widetilde{\mathcal{O}}(\sqrt{1+Q_T})$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Dynamic NE-Reg	$\widetilde{\mathcal{O}}\big(\min\{\sqrt{(1+V_T)(1+P_T)}+P_T, 1+W_T\}\big)$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Duality Gap	$\widetilde{\mathcal{O}}\big(\min\{T^{\frac{3}{4}}(1+Q_T)^{\frac{1}{4}}, T^{\frac{1}{2}}(1+Q_T^{\frac{3}{2}}+P_TQ_T)^{\frac{1}{2}}\}\big)$	

- $Q_T = V_T + \min\{P_T, W_T\}$

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Measure	Time-Varying Game	Fixed Game
Individual Regret	$\widetilde{\mathcal{O}}(\sqrt{1+Q_T})$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Dynamic NE-Reg	$\widetilde{\mathcal{O}}\big(\min\{\sqrt{(1+V_T)(1+P_T)}+P_T, 1+W_T\}\big)$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Duality Gap	$\widetilde{\mathcal{O}}\big(\min\{T^{\frac{3}{4}}(1+Q_T)^{\frac{1}{4}}, T^{\frac{1}{2}}(1+Q_T^{\frac{3}{2}}+P_TQ_T)^{\frac{1}{2}}\}\big)$	$\widetilde{\mathcal{O}}(\sqrt{T})$ recovers [WLZL'21]

- $Q_T = V_T + \min\{P_T, W_T\}$

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Measure	Time-Varying Game	Fixed Game
Individual Regret	$\widetilde{\mathcal{O}}(\sqrt{1+Q_T})$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Dynamic NE-Reg	$\widetilde{\mathcal{O}}\big(\min\{\sqrt{(1+V_T)(1+P_T)}+P_T, 1+W_T\}\big)$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Duality Gap	$\widetilde{\mathcal{O}}\big(\min\{T^{\frac{3}{4}}(1+Q_T)^{\frac{1}{4}}, T^{\frac{1}{2}}(1+Q_T^{\frac{3}{2}}+P_TQ_T)^{\frac{1}{2}}\}\big)$	$\widetilde{\mathcal{O}}(\sqrt{T})$ recovers [WLZL'21]

- $Q_T = V_T + \min\{P_T, W_T\}$
- the last column also holds when A_t changes $\mathcal{O}(1)$ times

Result Overview

We propose a parameter-free algorithm that obtains the following simultaneously (when deployed by both players):

Measure	Time-Varying Game	Fixed Game
Individual Regret	$\widetilde{\mathcal{O}}(\sqrt{1+Q_T})$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Dynamic NE-Reg	$\widetilde{\mathcal{O}}\big(\min\{\sqrt{(1+V_T)(1+P_T)}+P_T, 1+W_T\}\big)$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Duality Gap	$\widetilde{\mathcal{O}}\big(\min\{T^{\frac{3}{4}}(1+Q_T)^{\frac{1}{4}}, T^{\frac{1}{2}}(1+Q_T^{\frac{3}{2}}+P_TQ_T)^{\frac{1}{2}}\}\big)$	$\widetilde{\mathcal{O}}(\sqrt{T})$ recovers [WLZL'21]

- $Q_T = V_T + \min\{P_T, W_T\}$
- the last column also holds when A_t changes $\mathcal{O}(1)$ times
- robustness: $\mathrm{Reg}_T^x = \widetilde{\mathcal{O}}(\sqrt{T})$ even if y-player behaves arbitrarily

Algorithm Design: Review of RVU for a Fixed Game

For a fixed game ($A_t = A$), Syrgkanis-Agarwal-Luo-Schapire'15 proposes the "**Regret bounded by Variation in Utilities** (RVU)" property:

$$\mathrm{Reg}_T^x \leq \frac{\alpha}{\eta} + \eta\beta \sum_{t=2}^{T} \|Ay_t - Ay_{t-1}\|_\infty^2 - \frac{\gamma}{\eta} \sum_{t=2}^{T} \|x_t - x_{t-1}\|_1^2$$

for a learning rate parameter η and constants α, β, γ.

Algorithm Design: Review of RVU for a Fixed Game

For a fixed game ($A_t = A$), Syrgkanis-Agarwal-Luo-Schapire'15 proposes the "**Regret bounded by Variation in Utilities** (RVU)" property:

$$\mathrm{Reg}_T^x \leq \frac{\alpha}{\eta} + \eta\beta \sum_{t=2}^{T} \|Ay_t - Ay_{t-1}\|_\infty^2 - \frac{\gamma}{\eta} \sum_{t=2}^{T} \|x_t - x_{t-1}\|_1^2$$

for a learning rate parameter η and constants α, β, γ.

This is satisfied by many "optimistic" online learning algorithms (e.g. optimistic Hedge and optimistic GD).

Algorithm Design: Review of RVU for a Fixed Game

For a fixed game ($A_t = A$), Syrgkanis-Agarwal-Luo-Schapire'15 proposes the "**Regret bounded by Variation in Utilities** (RVU)" property:

$$\text{Reg}_T^x \leq \frac{\alpha}{\eta} + \eta\beta \sum_{t=2}^{T} \|Ay_t - Ay_{t-1}\|_\infty^2 - \frac{\gamma}{\eta} \sum_{t=2}^{T} \|x_t - x_{t-1}\|_1^2$$

for a learning rate parameter η and constants α, β, γ.

This is satisfied by many "optimistic" online learning algorithms (e.g. optimistic Hedge and optimistic GD).

Useful because $\|Ay_t - Ay_{t-1}\|_\infty^2 \leq \|y_t - y_{t-1}\|_1^2$, so $\text{Reg}_T^x + \text{Reg}_T^y$ is small.

Algorithm Design: Dynamic RVU

For time-varying games, it is important to compete with arbitrary time-varying strategies $u_1, \ldots, u_T \in \Delta_m$.

Algorithm Design: Dynamic RVU

For time-varying games, it is important to compete with arbitrary time-varying strategies $u_1, \ldots, u_T \in \Delta_m$. We thus propose Dynamic RVU:

$$\sum_{t=1}^{T}(x_t - u_t)^\top A_t y_t \leq \frac{\alpha P_T^u}{\eta} + \eta\beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta} \sum_{t=2}^{T} \|x_t - x_{t-1}\|_1^2$$

where $P_T^u = 1 + \sum_{t=2}^{T} \|u_t - u_{t-1}\|_1$.

Algorithm Design: Dynamic RVU

For time-varying games, it is important to compete with arbitrary time-varying strategies $u_1, \ldots, u_T \in \Delta_m$. We thus propose Dynamic RVU:

$$\sum_{t=1}^{T}(x_t - u_t)^\top A_t y_t \le \frac{\alpha P_T^u}{\eta} + \eta\beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta} \sum_{t=2}^{T} \|x_t - x_{t-1}\|_1^2$$

where $P_T^u = 1 + \sum_{t=2}^{T} \|u_t - u_{t-1}\|_1$.

Again, this is satisfied by standard algorithms such as:

- **optimistic GD**:

$$\widehat{x}_{t+1} = \operatorname*{argmin}_{x \in \Delta_m} \eta \langle x, A_t y_t \rangle + \|x - \widehat{x}_t\|^2$$

$$x_{t+1} = \operatorname*{argmin}_{x \in \Delta_m} \eta \langle x, A_t y_t \rangle + \|x - \widehat{x}_{t+1}\|^2$$

- **optimistic Hedge** over a truncated simplex

Combining Base Algorithms with DRVU

For each performance measure, DRVU implies a favorable guarantee, but requires a different tuning of η and the knowledge of V_T, W_T, P_T.

Combining Base Algorithms with DRVU

For each performance measure, DRVU implies a favorable guarantee, but requires a different tuning of η and the knowledge of V_T, W_T, P_T.

To achieve this for all measures simultaneously without knowing V_T, W_T, P_T, we propose to learn over a set of base algorithms with DRVU and different tunings, via another optimistic meta-algorithm.

Combining Base Algorithms with DRVU

For each performance measure, DRVU implies a favorable guarantee, but requires a different tuning of η and the knowledge of V_T, W_T, P_T.

To achieve this for all measures simultaneously without knowing V_T, W_T, P_T, we propose to learn over a set of base algorithms with DRVU and different tunings, via another optimistic meta-algorithm.

While standard, the right execution here requires two ideas.

Algorithm Overview (from x-player's perspective)

Input: any base algorithm $\mathcal{A}(\eta)$ satisfying DRVU with learning rate η.

Algorithm Overview (from x-player's perspective)

Input: any base algorithm $\mathcal{A}(\eta)$ satisfying DRVU with learning rate η.

Initialize: a set of $\mathcal{O}(\log T)$ base learners $\mathcal{S}$, each of which is $\mathcal{A}(\eta)$ with a certain η

Algorithm Overview (from x-player's perspective)

Input: any base algorithm $\mathcal{A}(\eta)$ satisfying DRVU with learning rate η.

Initialize: a set of $\mathcal{O}(\log T)$ base learners $\mathcal{S}$, each of which is $\mathcal{A}(\eta)$ with a certain η

For $t = 1, \ldots, T$:

- receive $x_{t,i} \in \Delta_m$ from each base learner $i \in \mathcal{S}$.

Algorithm Overview (from x-player's perspective)

Input: any base algorithm $\mathcal{A}(\eta)$ satisfying DRVU with learning rate η.

Initialize: a set of $\mathcal{O}(\log T)$ base learners $\mathcal{S}$, each of which is $\mathcal{A}(\eta)$ with a certain η

For $t = 1, \ldots, T$:

- receive $x_{t,i} \in \Delta_m$ from each base learner $i \in \mathcal{S}$.
- compute "prediction vector m_t" and update $p_t \in \Delta_{\mathcal{S}}$ as:

$$p_t = \operatorname*{argmin}_{p \in \Delta_{\mathcal{S}}} \epsilon_t \langle p, m_t \rangle + \|p - \widehat{p}_t\|_2^2$$

Algorithm Overview (from x-player's perspective)

Input: any base algorithm $\mathcal{A}(\eta)$ satisfying DRVU with learning rate η.

Initialize: a set of $\mathcal{O}(\log T)$ base learners $\mathcal{S}$, each of which is $\mathcal{A}(\eta)$ with a certain η

For $t = 1, \ldots, T$:

- receive $x_{t,i} \in \Delta_m$ from each base learner $i \in \mathcal{S}$.
- compute "prediction vector m_t" and update $p_t \in \Delta_{\mathcal{S}}$ as:

$$p_t = \operatorname*{argmin}_{p \in \Delta_{\mathcal{S}}} \epsilon_t \langle p, m_t \rangle + \|p - \widehat{p}_t\|_2^2$$

- play the final action $x_t = \sum_{i \in \mathcal{S}} p_{t,i} x_{t,i}$

Algorithm Overview (from x-player's perspective)

Input: any base algorithm $\mathcal{A}(\eta)$ satisfying DRVU with learning rate η.

Initialize: a set of $\mathcal{O}(\log T)$ base learners $\mathcal{S}$, each of which is $\mathcal{A}(\eta)$ with a certain η

For $t = 1, \ldots, T$:

- receive $x_{t,i} \in \Delta_m$ from each base learner $i \in \mathcal{S}$.
- compute "prediction vector m_t" and update $p_t \in \Delta_{\mathcal{S}}$ as:

$$p_t = \operatorname*{argmin}_{p \in \Delta_{\mathcal{S}}} \epsilon_t \langle p, m_t \rangle + \|p - \widehat{p}_t\|_2^2$$

- play the final action $x_t = \sum_{i \in \mathcal{S}} p_{t,i} x_{t,i}$
- suffer loss $x_t^\top A_t y_t$, observe $A_t y_t$, and send it to each base learner

Algorithm Overview (from x-player's perspective)

Input: any base algorithm $\mathcal{A}(\eta)$ satisfying DRVU with learning rate η.

Initialize: a set of $\mathcal{O}(\log T)$ base learners $\mathcal{S}$, each of which is $\mathcal{A}(\eta)$ with a certain η

For $t = 1, \ldots, T$:

- receive $x_{t,i} \in \Delta_m$ from each base learner $i \in \mathcal{S}$.
- compute "prediction vector m_t" and update $p_t \in \Delta_{\mathcal{S}}$ as:

$$p_t = \operatorname*{argmin}_{p \in \Delta_{\mathcal{S}}} \epsilon_t \langle p, m_t \rangle + \|p - \widehat{p}_t\|_2^2$$

- play the final action $x_t = \sum_{i \in \mathcal{S}} p_{t,i} x_{t,i}$
- suffer loss $x_t^\top A_t y_t$, observe $A_t y_t$, and send it to each base learner
- compute "loss vector ℓ_t" and update $\widehat{p}_{t+1}$ as:

$$\widehat{p}_{t+1} = \operatorname*{argmin}_{p \in \Delta_{\mathcal{S}}} \epsilon_t \langle p, \ell_t \rangle + \|p - \widehat{p}_t\|_2^2$$

Algorithm Overview (from x-player's perspective)

Input: any base algorithm $\mathcal{A}(\eta)$ satisfying DRVU with learning rate η.

Initialize: a set of $\mathcal{O}(\log T)$ base learners $\mathcal{S}$, each of which is $\mathcal{A}(\eta)$ with a certain η or a dummy learner always selecting a fixed action

Idea 1: make sure the meta-alg comparable to Nash

For $t = 1, \ldots, T$:

- receive $x_{t,i} \in \Delta_m$ from each base learner $i \in \mathcal{S}$
- compute "prediction vector m_t" and update

$$p_t = \operatorname*{argmin}_{p \in \Delta_\mathcal{S}} \epsilon_t \langle p, m_t \rangle + \|p - \widehat{p}_t\|_2^2$$

- play the final action $x_t = \sum_{i \in \mathcal{S}} p_{t,i} x_{t,i}$
- suffer loss $x_t^\top A_t y_t$, observe $A_t y_t$, and send it to each base learner
- compute "loss vector ℓ_t" and update $\widehat{p}_{t+1}$ as:

$$\widehat{p}_{t+1} = \operatorname*{argmin}_{p \in \Delta_\mathcal{S}} \epsilon_t \langle p, \ell_t \rangle + \|p - \widehat{p}_t\|_2^2$$

Algorithm Overview (from x-player's perspective)

Input: any base algorithm $\mathcal{A}(\eta)$ satisfying DRVU with learning rate η.

Initialize: a set of $\mathcal{O}(\log T)$ base learners $\mathcal{S}$, each of which is $\mathcal{A}(\eta)$ with a certain η or a dummy learner always selecting a fixed action

For $t = 1, \ldots, T$:

- receive $x_{t,i} \in \Delta_m$ from each base learner $i \in \mathcal{S}$
- compute "prediction vector m_t" and update

$$p_t = \operatorname*{argmin}_{p \in \Delta_{\mathcal{S}}} \epsilon_t \langle p, m_t \rangle + \|p - \widehat{p}_t\|_2^2$$

- play the final action $x_t = \sum_{i \in \mathcal{S}} p_{t,i} x_{t,i}$
- suffer loss $x_t^\top A_t y_t$, observe $A_t y_t$, and send it t
- compute "loss vector ℓ_t" and update $\widehat{p}_{t+1}$ as

$$\widehat{p}_{t+1} = \operatorname*{argmin}_{p \in \Delta_{\mathcal{S}}} \epsilon_t \langle p, \ell_t \rangle + \|p - \widehat{p}_t\|_2^2$$

Idea 1: make sure the meta-alg comparable to Nash

Idea 2: bias towards more stable learners

Explaining Idea 2

DRVU bound of base learner $i^\star$ with ideal tuning $\eta^\star$:

$$\frac{\alpha P_T^u}{\eta^\star} + \eta^\star \beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta^\star} \sum_{t=2}^{T} \underbrace{\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2}_{(*)}$$

Explaining Idea 2

DRVU bound of base learner $i^\star$ with ideal tuning $\eta^\star$:

$$\frac{\alpha P_T^u}{\eta^\star} + \eta^\star \beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta^\star} \sum_{t=2}^{T} \underbrace{\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2}_{(*)}$$

- $\|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2$ is related to $\|y_t - y_{t-1}\|_1^2$, not $\|y_{t,i^\star} - y_{t-1,i^\star}\|_1^2$

Explaining Idea 2

DRVU bound of base learner $i^\star$ with ideal tuning $\eta^\star$:

$$\frac{\alpha P_T^u}{\eta^\star} + \eta^\star \beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta^\star} \sum_{t=2}^{T} \underbrace{\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2}_{(*)}$$

- $\|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2$ is related to $\|y_t - y_{t-1}\|_1^2$, not $\|y_{t,i^\star} - y_{t-1,i^\star}\|_1^2$
- resolved by biasing towards stable learners;

Explaining Idea 2

DRVU bound of base learner $i^\star$ with ideal tuning $\eta^\star$:

$$\frac{\alpha P_T^u}{\eta^\star} + \eta^\star \beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta^\star} \sum_{t=2}^{T} \underbrace{\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2}_{(*)}$$

- $\|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2$ is related to $\|y_t - y_{t-1}\|_1^2$, not $\|y_{t,i^\star} - y_{t-1,i^\star}\|_1^2$
- resolved by biasing towards stable learners; define (similarly for m_t)

$$\ell_{t,i} = x_{t,i}^\top A_t y_t + \lambda \|x_{t,i} - x_{t-1,i}\|_1^2$$

Explaining Idea 2

DRVU bound of base learner $i^\star$ with ideal tuning $\eta^\star$:

$$\frac{\alpha P_T^u}{\eta^\star} + \eta^\star \beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta^\star} \sum_{t=2}^{T} \underbrace{\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2}_{(*)}$$

- $\|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2$ is related to $\|y_t - y_{t-1}\|_1^2$, not $\|y_{t,i^\star} - y_{t-1,i^\star}\|_1^2$
- resolved by biasing towards stable learners; define (similarly for m_t)

$$\ell_{t,i} = x_{t,i}^\top A_t y_t + \lambda \|x_{t,i} - x_{t-1,i}\|_1^2$$

- effect for the analysis:
  - introduce a positive term $\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2$, canceled by $(*)$

Explaining Idea 2

DRVU bound of base learner $i^\star$ with ideal tuning $\eta^\star$:

$$\frac{\alpha P_T^u}{\eta^\star} + \eta^\star \beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta^\star} \sum_{t=2}^{T} \underbrace{\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2}_{(*)}$$

- $\|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2$ is related to $\|y_t - y_{t-1}\|_1^2$, not $\|y_{t,i^\star} - y_{t-1,i^\star}\|_1^2$
- resolved by biasing towards stable learners; define (similarly for m_t)

$$\ell_{t,i} = x_{t,i}^\top A_t y_t + \lambda \|x_{t,i} - x_{t-1,i}\|_1^2$$

- effect for the analysis:
  - introduce a positive term $\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2$, canceled by $(*)$
  - introduce a negative term $-\sum_{i \in \mathcal{S}} p_{t,i} \|x_{t,i} - x_{t-1,i}\|_1^2$

Explaining Idea 2

DRVU bound of base learner $i^\star$ with ideal tuning $\eta^\star$:

$$\frac{\alpha P_T^u}{\eta^\star} + \eta^\star \beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta^\star} \sum_{t=2}^{T} \underbrace{\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2}_{(*)}$$

- $\|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2$ is related to $\|y_t - y_{t-1}\|_1^2$, not $\|y_{t,i^\star} - y_{t-1,i^\star}\|_1^2$
- resolved by biasing towards stable learners; define (similarly for m_t)

$$\ell_{t,i} = x_{t,i}^\top A_t y_t + \lambda \|x_{t,i} - x_{t-1,i}\|_1^2$$

- effect for the analysis:
  - introduce a positive term $\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2$, canceled by $(*)$
  - introduce a negative term $-\sum_{i \in \mathcal{S}} p_{t,i} \|x_{t,i} - x_{t-1,i}\|_1^2$
  - meta-algorithm itself satisfies RVU, providing a term $-\|p_t - p_{t-1}\|_1^2$

Explaining Idea 2

DRVU bound of base learner $i^\star$ with ideal tuning $\eta^\star$:

$$\frac{\alpha P_T^u}{\eta^\star} + \eta^\star \beta \sum_{t=2}^{T} \|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2 - \frac{\gamma}{\eta^\star} \sum_{t=2}^{T} \underbrace{\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2}_{(*)}$$

- $\|A_t y_t - A_{t-1} y_{t-1}\|_\infty^2$ is related to $\|y_t - y_{t-1}\|_1^2$, not $\|y_{t,i^\star} - y_{t-1,i^\star}\|_1^2$
- resolved by biasing towards stable learners; define (similarly for m_t)

$$\ell_{t,i} = x_{t,i}^\top A_t y_t + \lambda \|x_{t,i} - x_{t-1,i}\|_1^2$$

- effect for the analysis:
  - introduce a positive term $\|x_{t,i^\star} - x_{t-1,i^\star}\|_1^2$, canceled by $(*)$
  - introduce a negative term $-\sum_{i\in\mathcal{S}} p_{t,i} \|x_{t,i} - x_{t-1,i}\|_1^2$
  - meta-algorithm itself satisfies RVU, providing a term $-\|p_t - p_{t-1}\|_1^2$
  - the last two negative terms together cancel $\|x_t - x_{t-1}\|_1^2$

Key Stability Lemma

The two ideas enable us to prove the following key lemma, critical for bounding all three measures:

Key Stability Lemma

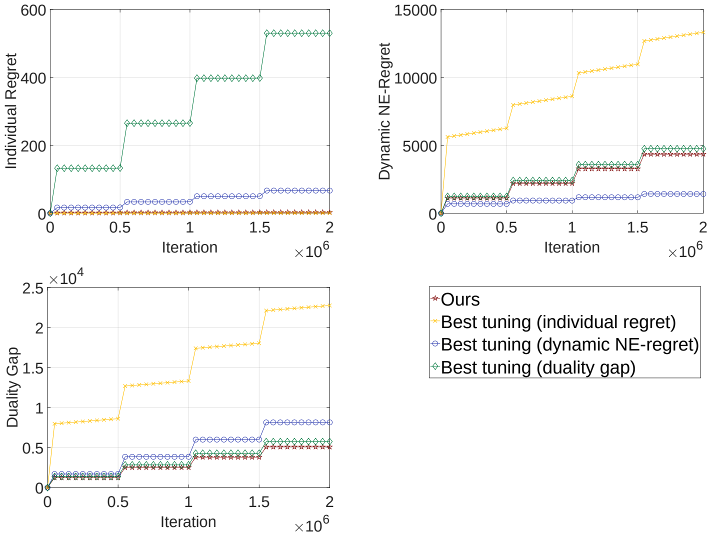
The two ideas enable us to prove the following key lemma, critical for bounding all three measures:

When deployed by both players, our algorithm ensures

$$\max\left\{\sum_{t=2}^{T}\|x_t - x_{t-1}\|_1^2,\ \sum_{t=2}^{T}\|y_t - y_{t-1}\|_1^2\right\}$$
$$= \widetilde{\mathcal{O}}\left(\min\left\{\sqrt{(1+V_T)(1+P_T)} + P_T, 1 + W_T\right\}\right)$$

Experiments

A synthetic environment s.t. $P_T = \Theta(\sqrt{T})$, $W_T = \Theta(T^{\frac{3}{4}})$, $V_T = \Theta(\sqrt{T})$.



Summary

Takeaway:

- how to measure performance when learning over time-varying games requires more thoughts;

Summary

Takeaway:

- how to measure performance when learning over time-varying games requires more thoughts;
- simultaneous strong guarantees under different measures are possible.

Summary

Takeaway:

- how to measure performance when learning over time-varying games requires more thoughts;
- simultaneous strong guarantees under different measures are possible.

Measure	Time-Varying Game	Fixed Game
Individual Regret	$\widetilde{\mathcal{O}}\big(\sqrt{1+V_T+\min\{P_T,W_T\}}\big)$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Dynamic NE-Reg	$\widetilde{\mathcal{O}}\big(\min\{\sqrt{(1+V_T)(1+P_T)}+P_T, 1+W_T\}\big)$	$\widetilde{\mathcal{O}}(1)$ recovers [HAM'21]
Duality Gap	$\widetilde{\mathcal{O}}\big(\min\{T^{\frac{3}{4}}(1+Q_T)^{\frac{1}{4}}, T^{\frac{1}{2}}(1+Q_T^{\frac{3}{2}}+P_TQ_T)^{\frac{1}{2}}\}\big)$	$\widetilde{\mathcal{O}}(\sqrt{T})$ recovers [WLZL'21]

Open Questions

1. We have considered full-info mixture feedback (x-player sees $A_t y_t$).

Open Questions

1. We have considered full-info mixture feedback (x-player sees $A_t y_t$).
   - bandit mixture feedback? (x-player sees $e_{i_t}^\top A_t y_t$ where $i_t \sim x_t$)

Open Questions

1. We have considered full-info mixture feedback (x-player sees $A_t y_t$).
   - bandit mixture feedback? (x-player sees $e_{i_t}^\top A_t y_t$ where $i_t \sim x_t$)
   - fully realization-based feedback? (x-player sees $e_{i_t}^\top A_t e_{j_t}$ where $i_t \sim x_t$ and $j_t \sim y_t$)

Open Questions

1. We have considered full-info mixture feedback (x-player sees $A_t y_t$).
   - bandit mixture feedback? (x-player sees $e_{i_t}^\top A_t y_t$ where $i_t \sim x_t$)
   - fully realization-based feedback? (x-player sees $e_{i_t}^\top A_t e_{j_t}$ where $i_t \sim x_t$ and $j_t \sim y_t$)
   - many techniques in this work fail :(

Open Questions

1. We have considered full-info mixture feedback (x-player sees $A_t y_t$).
   - bandit mixture feedback? (x-player sees $e_{i_t}^\top A_t y_t$ where $i_t \sim x_t$)
   - fully realization-based feedback? (x-player sees $e_{i_t}^\top A_t e_{j_t}$ where $i_t \sim x_t$ and $j_t \sim y_t$)
   - many techniques in this work fail :(
2. Time-varying general-sum games?

Open Questions

1. We have considered full-info mixture feedback (x-player sees $A_t y_t$).
   - bandit mixture feedback? (x-player sees $e_{i_t}^\top A_t y_t$ where $i_t \sim x_t$)
   - fully realization-based feedback? (x-player sees $e_{i_t}^\top A_t e_{j_t}$ where $i_t \sim x_t$ and $j_t \sim y_t$)
   - many techniques in this work fail :(
2. Time-varying general-sum games?
3. Recall: for time-varying games
   - changes could depend on players' actions: e.g. Markov games
   - changes could also only caused by exogenous factors

Open Questions

1. We have considered full-info mixture feedback (x-player sees $A_t y_t$).
   - bandit mixture feedback? (x-player sees $e_{i_t}^\top A_t y_t$ where $i_t \sim x_t$)
   - fully realization-based feedback? (x-player sees $e_{i_t}^\top A_t e_{j_t}$ where $i_t \sim x_t$ and $j_t \sim y_t$)
   - many techniques in this work fail :(
2. Time-varying general-sum games?
3. Recall: for time-varying games
   - changes could depend on players' actions: e.g. Markov games
   - changes could also only caused by exogenous factors
   - or, changes could come from both! (time-varying Markov games)